

RAW SEQUENCE LISTING

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) no errors detected.

Application Serial Number: 10/518, 019A
Source: Pg 7/10
Date Processed by STIC: 7/5/05

ENTERED



PCT

RAW SEQUENCE LISTING

DATE: 07/05/2005

PATENT APPLICATION: US/10/518,019A

TIME: 11:10:11

Input Set : A:\PTO.RJ.txt

Output Set: N:\CRF4\07052005\J518019A.raw

3 <110> APPLICANT: Commissariat a l'Energie Atomique
 4 Centre National de la Recherche Scientifique
 5 GONDRY, Muriel
 6 GENET, Roger
 7 LAUTRU, Sylvie
 8 PERNODET, Jean-Luc
 10 <120> TITLE OF INVENTION: Polynucleotides and polypeptides coded by said
 polynucleotides

11 involved in the synthesis of diketopiperazine derivatives

13 <130> FILE REFERENCE: CGA263/83FR

15 <140> CURRENT APPLICATION NUMBER: US/10/518,019A

16 <141> CURRENT FILING DATE: 2004-12-15

18 <160> NUMBER OF SEQ ID NOS: 23

20 <170> SOFTWARE: PatentIn version 3.3

22 <210> SEQ ID NO: 1

23 <211> LENGTH: 657

24 <212> TYPE: DNA

25 <213> ORGANISM: Streptomyces noursei

27 <400> SEQUENCE: 1

28 gtgaggcgcc acccatcgca ttcgccgtac cgcgggcgggt gtgaggtgcg cccaaaaaga 60
 30 aggggattga tgtagctca cagttcatct gaatcgccgc cggaatcctt gccggacgcg 120
 32 tggacgggcc tcaaaacccg taccgccgtc cgcaattacg cgaaagagcc ggtcgacgac 180
 34 gcgctgatcg agcagctggt ggaggccatg ctccgcgcgc cgaccgcctc caaccggcag 240
 36 gcgtgggtcgt tcatggtggt gcgcaggccc gccgcggctc gccgtttctcg 300
 38 cccgggggtgc tgggaacccc cgccttcttc gtcgtggcct gcgtcgaccg cagtctgacc 360
 40 gacaacctct cccgaagct ctgcgagaag atctacgaca ccagcaagct ctgtgtcgcc 420
 42 atggcggtgg agaacctgct gctcgcgccg cacgcggccg gcctgggagg atgcccggtg 480
 44 ggcagcttca ggtccgacat cgtcaccagc atgctcggta tcccgaaca catcgagccg 540
 46 atgctcgtgg tcccgatcgg ccgtcccgcg acagccctcg tcccctcca gcgccgcgcc 600
 48 aagaatgagg tcgtcaacta tgaatcctgg ggaaaccgtg ctgccgcccc aactgcyg 657

51 <210> SEQ ID NO: 2

52 <211> LENGTH: 318

53 <212> TYPE: DNA

54 <213> ORGANISM: Streptomyces noursei

56 <400> SEQUENCE: 2

57 atgaatcctg gggaaaccgt gctgccgccc caactgcgtg aggagatcgc gtcctcgcc 60
 59 gtctatctgc tcagcagcgg ccgcggactc ctggaggagc cggccgacta cggaatttac 120
 61 cgctgtaccg acggggcccg tcgggcgctc caactcctcg acgaacacgg cgggagcacg 180
 63 gcacggctga ccgccgtccg cgagcgtctc gacgaggtca tggtcgcgcc gatgggcgag 240
 65 gaccgggaca tgggcgcgat tctggacgac ctgtgtcgcc aaatggcaga cgctcttcgg 300
 67 gaaattgaaa ccccctga 318

70 <210> SEQ ID NO: 3

71 <211> LENGTH: 720

72 <212> TYPE: DNA

RAW SEQUENCE LISTING

DATE: 07/05/2005

PATENT APPLICATION: US/10/518,019A

TIME: 11:10:11

Input Set : A:\PTO.RJ.txt

Output Set: N:\CRF4\07052005\J518019A.raw

73 <213> ORGANISM: Streptomyces noursei

75 <400> SEQUENCE: 3

```

76 atgcttgacg gcttagttcc cgcgcgggac caggaatgc gggaagaaat acttggcgac 60
77 cgcagccgat tgatccggca acgcggtgag cagccctca tcggaatcag tgcgggcaac 120
78 agttatttca gccagaagaa caccgtcatg ctgctgcaat gggccgggca gcgtttcgag 180
79 cgcaccgatg tegtctatgt cgacaccac atcgacgaga tgctgatcgc cgacggccgc 240
80 agcgcgcagg aggccgagcg gtcggtcaaa cgcacgctca aggatctgcg gcgcagactc 300
81 cggcgctcgc tggagagcgt gggcgaccac gccgagcggt tccgtgtccg gtccctgtcc 360
82 gagctccagg agaccctga gtaccgggcc gtacgcgagc gcaccgaccg ggccctcgag 420
83 gaggacgccg aattcgccac gcctgcgag gacatggtgc gggccgtggt gatgaaccgg 480
84 cccggtgacg gcgtcgcat ctcccgaggaa cacctgcggg ccggtctgaa ctacgtgctg 540
85 gccgaggccc cgtcttcgc ggaactcgccc ggagtcttct ccgtccctc ctcggtgctc 600
86 tgctaccaca tcgacacccc gatcacggcg ttcctgtccc ggcgcgagac cggtttccgg 660
87 gcggccgagg gacaggcgta cgtcgtcgtc agggccagg agctggccga cgcggcctag 720

```

101 <210> SEQ ID NO: 4

102 <211> LENGTH: 834

103 <212> TYPE: DNA

104 <213> ORGANISM: Streptomyces noursei

106 <400> SEQUENCE: 4

```

107 atgtcatggg gaggacagga cacttgctca tggtgcgga cggggccctt cggcgaagct 60
108 gaagacgtag gaagacagca cagtcgcac gccgggggac ccgtcatgac tcaagccgcc 120
109 accgtcaccg ccaccacgag ccaggggcagg gcactcctgc ggagcctgac gccgtgttc 180
110 gtggacgccg cgatcccgtc cggctcgtac ttctcctcg ccgagggtt cggcatgagc 240
111 acggtcgccg cgctggcctg gagcagcgtg gtcccggcgc tgcgcacgat ctggggcctg 300
112 gtccgggagc ggacgggtcaa cggcctcgcg ctgctgatcc tcgtcgtcaa cgtggtgggg 360
113 ctggcgacga gcaccctgac cggcgatgcc cggctgatga tggccaagga cagcggcgctc 420
114 agcagcgctc tcgggatcgc gatcctgctc tcggtgcgcg gccggcgccc gctgatgacc 480
115 gccggactcc ggccctgggt gaccaaggga agcccggagg ggaacgccgc atgggaccgg 540
116 ctgtgggcgc gcagcgcgcg gttccggcaa ctggagcggc gattctcgac ggtctggggg 600
117 agcgcctcgc tgatcgagtg cgtggtcaa gtcgtcggtg cgtacgtcct gccggtgcac 660
118 accatggtgt ggctgggcac ggtgctgacg gtggtggcga tctgtctggc catggtggtc 720
119 gcgggcggcg gcagcgccga gccgatggag cggatggtca agggcgaggc cggggccgcc 780
120 ggcgaggccg ccacggcggg gaacggcgag ccggcgccgg ccgccgcggc ctga 834

```

136 <210> SEQ ID NO: 5

137 <211> LENGTH: 3839

138 <212> TYPE: DNA

139 <213> ORGANISM: Streptomyces noursei

141 <400> SEQUENCE: 5

```

142 ggatccgtcc cgacggggcg gaaccgggtga ggcgccacc atcgcatcgc ccgtaccgcg 60
143 gcgggtgtga ggtgcgccc aaaagaagg gattgatgtt agctcacagt tcatctgaat 120
144 cgccgccgga atccttgccg gacgcgtgga cggctcctaa aaccgtacc gccgtccgca 180
145 attacgcgaa agagccgggtc gacgacgcgc tgatcgagca gctgttgag gccatgctcg 240
146 ccgcgccgac cgctccaac cggcaggcgt ggtcgttcat ggtggtgcgc agggccgccg 300
147 cggtcgcccg gctgcgcgc ttctcgccc ggggtgctggg aacccccgcc ttcttcgtcg 360
148 tggcctgcgt cgaccgcagt ctgaccgaca acctctccc gaagctctcg cagaagatct 420
149 acgacaccag caagctctgt gtcgccatgg cgggtggagaa cctgctgctc gcggcgacg 480
150 cggccggcct gggcggatgc ccggtgggca gcttcaggct cgacatcgct accagcatgc 540
151 tcggtatccc ggaacacatc gagccgatgc tcgtgggtccc gatcgggcgt cccgcgacag 600
152 cctcgtcccc ctcccagcgc cgcgccaaga atgaggtcgt caactatgaa tcttggggaa 660

```

RAW SEQUENCE LISTING

DATE: 07/05/2005

PATENT APPLICATION: US/10/518,019A

TIME: 11:10:11

Input Set : A:\PTO.RJ.txt

Output Set: N:\CRF4\07052005\J518019A.raw

164	accgtgctgc	cgccccaaact	gcgtgaggag	atcgcgctcc	tcgccgtcta	tctgctcagc	720
166	agcggccgcg	gactcctgga	ggagccggcc	gactacggaa	tttaccgctg	taccgacggg	780
168	gcccgtcggg	cgctccaact	cctcgacgaa	cacggcggga	gcacggcacg	gctgaccgcc	840
170	gtccgcgagc	gtctcgacga	ggcatgttgc	gcgccgatgg	gcgaggaccg	ggacatgggc	900
172	gcgattctgg	acgacctgtg	tcgccaaatg	gcagacgctc	ttccggaaat	tgaaaccccc	960
174	tgacggctgt	ccggggcaac	cccaaaagga	cttcttagca	tgcttgacag	cttagttccc	1020
176	gcgccggacc	acggaatgcg	ggaagaaata	cttgccgacc	gcagccgatt	gatccggcaa	1080
178	cgccggtgagc	acgccctcat	cggaatcagt	gcgggcaaca	gttatttcag	ccagaagaac	1140
180	accgtcatgc	tgctgcaatg	ggccgggcag	cgtttcgagc	gcaccgatgt	cgtctatgtc	1200
182	gacacccaca	tcgacgagat	gctgatcgcc	gacggccgca	gcgcgcagga	ggccgagcgg	1260
184	tcggtcaaac	gcacgctcaa	ggatctgcgg	gcgagactcc	ggcgctcgct	ggagagcgtg	1320
186	ggcgaccacg	ccgagcgggt	ccgtgtcccg	tccctgtccg	agctccagga	gaccctgag	1380
188	taccgggccg	tacgcgagcg	caccgaccgg	gccttcgagg	aggacgccga	attcgccacc	1440
190	gcctgcgagg	acatggtgcg	ggccgtggtg	atgaaccggc	ccggtgacgg	cgtcggcatc	1500
192	tccgcggaac	acctgcgggc	cgtctgaac	tacgtgctgg	ccgaggcccc	gctcttcgcg	1560
194	gactcgcccc	gagctcttctc	cgtcccctcc	tcggtgctct	gctaccacat	cgacaccccc	1620
196	atcacggcgt	tcctgtcccc	gcgcgagacc	ggtttccggg	cggccgaggg	acaggcgtac	1680
198	gtcgtcgtea	ggccccagga	gctggccgac	gcggcctagt	tgggggcgtc	cgcgggcgga	1740
200	cctgcctccc	caccgcgtcc	cggtgccggc	gccgggcatg	acaaatgtca	tggggaggac	1800
202	aggacacttg	ctcatggtgc	ggaacggggc	ccctcggcga	agctgaagac	gtaggaagac	1860
204	agcacacgtc	gcacgccggg	ggaccctca	tgactcaagc	cgccaccgtc	accgccacca	1920
206	cgagccaggg	cagggcactc	ctgcggagcc	tgacgccgct	gttcgtggac	gccgcgatcc	1980
208	cgctcggctc	gtacttcctc	ctcgccgagg	gcttcggcat	gagcacggtc	gccgcgctgg	2040
210	cctggagcag	cgtggtcccg	gcgctgcgca	cgatctgggg	cctggtccgg	gagcggacgg	2100
212	tcaacggcct	cgcgctgctg	atcctcgctc	tcaacgtggt	ggggctggcg	acgagcacc	2160
214	tgaccggcga	tgcccggctg	atgatggcca	aggacagcgg	cgtcagcagc	gtcgtcggga	2220
216	tcgcgaccc	gctctcggtg	cgcgccgggc	gcccgtgat	gaccgcccga	ctccggccct	2280
218	gggtgaccaa	gggaagcccc	gaggggaacg	ccgcatggga	ccggctgtgg	gcgcgcagcg	2340
220	cgcggttccg	gcaactggag	cggcgattct	cgacggctctg	ggggagcgcc	ctgctgatcg	2400
222	agtgcgtggt	caaggctcgt	ggtgcgtacg	tcctgcgggt	gcacaccatg	gtgtggctgg	2460
224	gcacgggtgt	gacggtgggt	gcgatcctgc	tggccatggt	ggtcgcgggc	ggcgccagcg	2520
226	ccgagccgat	ggagccgatg	gtcaaggccg	aggctcgggg	cgccggcgag	gccgccacgg	2580
228	cggggaaacg	cgagccggcg	cgccgcggcg	cggcctgaga	ccgcgcggcg	ggggagttag	2640
230	ggaaatcgcg	tacaggattg	gcgcgtcgag	cacccccgcc	ctcgataggg	cggggccccg	2700
232	gcgcgatatg	tcggcgatgc	gacgggacat	cggagccccg	cgtcgacggt	tcaacggcga	2760
234	tccggacggc	acgcggcttt	cgtcggccac	gaagggaacg	gaagtcatgt	cgactgttca	2820
236	cactggggtc	acgcagagcg	gtctcaccgc	cgagctggcc	tccctgcacg	ccgagctcgt	2880
238	ccgtcggaat	cccgtgaaag	cggagtcca	ccaggcggcc	ctggaggtcc	tcgaaacgct	2940
240	ggcaccgggt	ctcaccgccc	ggccggagtt	cgccgacgcc	aaggctcctg	agcggatcgt	3000
242	cgagccggag	cggcagatca	tgttcgcgct	gccctggcag	gacgactccg	gcacgatccg	3060
244	ggtcaaccgc	ggcttcgggg	tggagtcaaa	cagcgcgctc	ggccccctaca	agggcggcct	3120
246	gcggttccac	gcgtccgtca	acctcggcat	cgtgaagttc	ctcggcttcg	agcagatctt	3180
248	caagaacgcc	ctgaccgggc	tgaacatcgg	cggcggcaag	ggcggcagcg	acttcgaccc	3240
250	gcacggcagg	tcggacgccg	aggtgatgcg	cttctgccag	tccttcatga	ccgagctgca	3300
252	ccgtcacctg	ggcgagcaca	ccgacgtgcc	ggctggcgac	atcggcgtcg	gcggccggga	3360
254	gatcggttac	ctcttcggcc	agtaccggcg	gatcaccaac	cgctgggagg	ccggcgctcc	3420
256	gaccggcaag	ggcctggcgt	ggggcggtct	caaggcccg	acggaggcca	ccggttacgg	3480
258	caatgtgctg	ttcaccgagg	agatgctcaa	gcagcgcgcc	gaggagctgg	acggccagca	3540
260	ggtggtggtc	tccgggtccg	gcaacgtcgc	catctacacc	atcgagaagg	cccaggcgct	3600

RAW SEQUENCE LISTING

DATE: 07/05/2005

PATENT APPLICATION: US/10/518,019A

TIME: 11:10:11

Input Set : A:\PTO.RJ.txt

Output Set: N:\CRF4\07052005\J518019A.raw

```

262 cggcgccaac gtcctgaccg tctcggactc cggcgggtac gtcgtcgacg agaagggcat 3660
264 cgacctggcg ctgctcaagc aggtcaagga ggctcgagcgc ggccgggtcg gcgactacgc 3720
266 ccagcggcgc ggcagttcgg cgaagtacgt cgccggcggg agcgtgtggg acgtcgctg 3780
268 tgacgtggcg ctgccgtcgg ccaccagaa cgagctcgac gcggacgccg cccggatcc 3839

```

271 <210> SEQ ID NO: 6

272 <211> LENGTH: 219

273 <212> TYPE: PRT

274 <213> ORGANISM: Streptomyces noursei

276 <400> SEQUENCE: 6

```

278 Met Arg Arg His Pro Ser His Ser Pro Tyr Arg Gly Gly Cys Glu Val
279 1 5 10 15
282 Arg Pro Lys Arg Arg Gly Leu Met Leu Ala His Ser Ser Ser Glu Ser
283 20 25 30
286 Pro Pro Glu Ser Leu Pro Asp Ala Trp Thr Val Leu Lys Thr Arg Thr
287 35 40 45
290 Ala Val Arg Asn Tyr Ala Lys Glu Pro Val Asp Asp Ala Leu Ile Glu
291 50 55 60
294 Gln Leu Leu Glu Ala Met Leu Ala Ala Pro Thr Ala Ser Asn Arg Gln
295 65 70 75 80
298 Ala Trp Ser Phe Met Val Val Arg Arg Pro Ala Ala Val Arg Arg Leu
299 85 90 95
302 Arg Ala Phe Ser Pro Gly Val Leu Gly Thr Pro Ala Phe Phe Val Val
303 100 105 110
306 Ala Cys Val Asp Arg Ser Leu Thr Asp Asn Leu Ser Pro Lys Leu Ser
307 115 120 125
310 Gln Lys Ile Tyr Asp Thr Ser Lys Leu Cys Val Ala Met Ala Val Glu
311 130 135 140
314 Asn Leu Leu Leu Ala Ala His Ala Ala Gly Leu Gly Gly Cys Pro Val
315 145 150 155 160
318 Gly Ser Phe Arg Ser Asp Ile Val Thr Ser Met Leu Gly Ile Pro Glu
319 165 170 175
322 His Ile Glu Pro Met Leu Val Val Pro Ile Gly Arg Pro Ala Thr Ala
323 180 185 190
326 Leu Val Pro Ser Gln Arg Arg Ala Lys Asn Glu Val Val Asn Tyr Glu
327 195 200 205
330 Ser Trp Gly Asn Arg Ala Ala Ala Pro Thr Ala
331 210 215

```

334 <210> SEQ ID NO: 7

335 <211> LENGTH: 104

336 <212> TYPE: PRT

337 <213> ORGANISM: Streptomyces noursei

339 <400> SEQUENCE: 7

```

341 Asn Pro Gly Glu Thr Val Leu Pro Pro Gln Leu Arg Glu Glu Ile Ala
342 1 5 10 15
345 Leu Leu Ala Val Tyr Leu Leu Ser Ser Gly Arg Gly Leu Leu Glu Glu
346 20 25 30
349 Pro Ala Asp Tyr Gly Ile Tyr Arg Cys Thr Asp Gly Ala Arg Arg Ala
350 35 40 45
353 Leu Gln Leu Leu Asp Glu His Gly Gly Ser Thr Ala Arg Leu Thr Ala

```

RAW SEQUENCE LISTING

DATE: 07/05/2005

PATENT APPLICATION: US/10/518,019A

TIME: 11:10:11

Input Set : A:\PTO.RJ.txt

Output Set: N:\CRF4\07052005\J518019A.raw

```

354      50      55      60
357 Val Arg Glu Arg Leu Asp Glu Val Met Phe Ala Pro Met Gly Glu Asp
358 65      70      75      80
361 Arg Asp Met Gly Ala Ile Leu Asp Asp Leu Cys Arg Gln Met Ala Asp
362      85      90      95
365 Ala Leu Pro Glu Ile Glu Thr Pro
366      100
369 <210> SEQ ID NO: 8
370 <211> LENGTH: 99
371 <212> TYPE: PRT
372 <213> ORGANISM: Streptomyces noursei
374 <400> SEQUENCE: 8
376 Met Leu Pro Pro Gln Leu Arg Glu Glu Ile Ala Leu Leu Ala Val Tyr
377 1      5      10      15
380 Leu Leu Ser Ser Gly Arg Gly Leu Leu Glu Glu Pro Ala Asp Tyr Gly
381      20      25      30
384 Ile Tyr Arg Cys Thr Asp Gly Ala Arg Arg Ala Leu Gln Leu Leu Asp
385      35      40      45
388 Glu His Gly Gly Ser Thr Ala Arg Leu Thr Ala Val Arg Glu Arg Leu
389      50      55      60
392 Asp Glu Val Met Phe Ala Pro Met Gly Glu Asp Arg Asp Met Gly Ala
393 65      70      75      80
396 Ile Leu Asp Asp Leu Cys Arg Gln Met Ala Asp Ala Leu Pro Glu Ile
397      85      90      95
400 Glu Thr Pro
404 <210> SEQ ID NO: 9
405 <211> LENGTH: 239
406 <212> TYPE: PRT
407 <213> ORGANISM: Streptomyces noursei
409 <400> SEQUENCE: 9
411 Met Leu Ala Gly Leu Val Pro Ala Pro Asp His Gly Met Arg Glu Glu
412 1      5      10      15
415 Ile Leu Gly Asp Arg Ser Arg Leu Ile Arg Gln Arg Gly Glu His Ala
416      20      25      30
419 Leu Ile Gly Ile Ser Ala Gly Asn Ser Tyr Phe Ser Gln Lys Asn Thr
420      35      40      45
423 Val Met Leu Leu Gln Trp Ala Gly Gln Arg Phe Glu Arg Thr Asp Val
424      50      55      60
427 Val Tyr Val Asp Thr His Ile Asp Glu Met Leu Ile Ala Asp Gly Arg
428 65      70      75      80
431 Ser Ala Gln Glu Ala Glu Arg Ser Val Lys Arg Thr Leu Lys Asp Leu
432      85      90      95
435 Arg Arg Arg Leu Arg Arg Ser Leu Glu Ser Val Gly Asp His Ala Glu
436      100      105      110
439 Arg Phe Arg Val Arg Ser Leu Ser Glu Leu Gln Glu Thr Pro Glu Tyr
440      115      120      125
443 Arg Ala Val Arg Glu Arg Thr Asp Arg Ala Phe Glu Asp Ala Glu
444      130      135      140
447 Phe Ala Thr Ala Cys Glu Asp Met Val Arg Ala Val Val Met Asn Arg

```

RAW SEQUENCE LISTING ERROR SUMMARY
PATENT APPLICATION: US/10/518,019A

DATE: 07/05/2005
TIME: 11:10:12

Input Set : A:\PTO.RJ.txt
Output Set: N:\CRF4\07052005\J518019A.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:12; Xaa Pos. 8

Seq#:13; Xaa Pos. 3

Invalid <213> Response:

Use of "Artificial" only as "<213> Organism" response is incomplete, per 1.823(b) of New Sequence Rules. Valid response is Artificial Sequence.

Seq#:14,15,16,17,18,19,20,21,22,23

VERIFICATION SUMMARY

DATE: 07/05/2005

PATENT APPLICATION: US/10/518,019A

TIME: 11:10:12

Input Set : A:\PTO.RJ.txt

Output Set: N:\CRF4\07052005\J518019A.raw

L:578 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:12 after pos.:0

L:595 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:13 after pos.:0